

Method of controlling devices via speech signals, more particularly, in motorcars

The invention relates to a method of controlling function units of a motorcar, or of devices installed in a motorcar, by means of speech signals. The invention also relates to a hardware configuration for implementing this method.

5

Basically, the method according to the invention permits to be used for any speech-controlled devices in which noise signal portions applied to the device depend on the state of operation and/or the operation environment of the respective device.

When function units of a motorcar (for example, windscreen wiper motor) and
10 of devices installed in a motorcar (for example, radio control, navigation system or mobile telephone) are controlled by means of speech signals to be recognized by a speech recognition system, noise signals are to be taken into account which depend on the operating state and/or operation environment of the motorcar to avoid faulty control of the function units or devices, respectively.

15 From JP 57-30913 (A) is known to detect by means of sensors both the speed of a motorcar and the gear shifted into. A reference voltage of a noise signal is generated from the sensor signals, which reference voltage indicates a measure for the current noise level (noise signal level) inside the motorcar. The noise signal reference voltage is compared to the output voltage of a speech input unit. When speech control signals are available, the
20 speech input unit receives acoustic signals which contain both noise signal portions and speech signal portions, which is reflected in the output voltage of the speech input unit. The output voltage of the speech input unit is compared to the noise signal reference voltage. If the output voltage of the speech input unit is higher than the noise signal reference voltage, a speech recognition system is activated. If the output voltage of the speech input unit drops to
25 below the noise signal reference voltage, the speech recognition system is deactivated.

From JP 6-83387 (A) is known to provide a vibration sensor in a motorcar to assess the vibration of the motorcar as a noise source. Furthermore, a first microphone is installed inside the motorcar to detect noise signals occurring in the interior of the motorcar. A second microphone in the interior of the motorcar is used for detecting speech signals

SCANNED, # 14

which are to be recognized by a speech recognition system. The second microphone receives acoustic signals, however, which contain noise signal portions in addition to the speech signal portions. With the aid of the signals of the vibration sensor, with the aid of the microphone signals of the first microphone and with the aid of two adaptive filters, the noise
5 signal level in the microphone signals generated by the second microphone is reduced; the thus generated signals with reduced noise signal portions are applied to a speech recognition system.

It is an object of the invention to effectively counteract noise signal effects by means of the method defined in the opening paragraph.

10 The object is achieved for applications to motorcars in that
acoustic signals occurring in the motorcar, which contain noise signal portions that depend on the operating state and/or operation environment of the motorcar and speech signal portions, as appropriate, are applied to a speech recognition system, and
the speech recognition system uses acoustic references which are selected
15 and/or adapted in dependence on detected data of the operating state and/or operation environment.

The method according to the invention is advantageous in that, based on data of operating states or operation environment which are often each easy to determine, the set for the automatic speech recognition is suitably adapted to acoustic references to be used. In a
20 motorcar, data about the operating state or operation environment can be read out from, for example, an on-board computer, which is connected to one or more detectors for determining the operating state or the operation environment of the motorcar. Starting from the determined operating state or operation environment, respectively, the noise signal portions are estimated indirectly. An extraction of the noise signal portions from the acoustic signals
25 fed to the speech recognition system can thus be made redundant. An estimation of the noise signal portions may be made so that predefined acoustic references are selected in dependence on the detected operating state and/or operation environment to model speech pauses in which the acoustic signals only have noise signal portions. In correspondence therewith, the presence of speech signal portions can be detected which is then the case when
30 there is no speech pause; in this manner an erroneous detection of the presence of speech signal portions when noise signal portions are changed can be avoided. The measures according to the invention enhance the reliability and safety of use of the whole system.

Also acoustic references representing speech signal portions can be adapted via the detected data of the operating state or operation environment, so that these superimposed noise signal portions are represented by the acoustic references.

5 An arrangement suitable for implementing the method according to the invention is stated in patent claim 8.

For optional speech-controlled devices, the object is achieved accordingly as defined in the characteristic features of patent claims 8 (method) and 9 (arrangement).

These and other aspects of the invention are apparent from and will be elucidated with reference to the embodiments described hereinafter.

10

In the drawings:

Fig. 1 shows the essential components for implementing the method according to the invention in a motorcar,

15

Fig. 2 shows a first possibility of generating an acoustic reference for a speech pause section and

Fig. 3 shows a second possibility of generating an acoustic reference for a speech pause section.

20

The block diagram shown in Fig. 1 describes the control of devices or function units in a motorcar. The devices/function units are represented here, for example, by the blocks 1a and 1b. The control takes place via speech signals which are fed via a microphone 2 to an automatic speech recognition system 3, whose recognition results are evaluated by a function unit 4 which causes a conversion into electric control signals to be supplied to the

25

A function unit 5 describes the extraction of features of microphone signals supplied by the microphone 2, where features for the individual successive signal sections are customarily combined to feature vectors. For the feature analysis, an acoustic signal is, for example, sampled, quantized and, finally, subjected to a cepstral analysis. Then there is a

30 subdivision of the acoustic signals into successive frames which partly overlap; for each frame a feature vector is formed. The feature vector components are formed by the determined cepstral values. Function block 6 describes customary procedures for comparison, in which the feature vectors are compared to an acoustic model 7 via customary search procedures, which lead to the speech recognition result applied to the function unit 4. The

comparison 6 and the acoustic model 7 are based on so-called Hidden Markov Models. The acoustic model 7 has acoustic references 8 and a lexicon 9. A respective acoustic reference is then assigned to a word sub-unit of one or more phonemes. The lexicon 9 defines associated sequences of word sub-units in accordance with the words combined in the lexicon.

5 The speech recognition system 3 includes an interface 10 which sets up a connection to an on-board computer 11 of the motorcar. The on-board computer 11 in its turn is connected by a connection 12a to at least one detector, which detects operating state data and/or operation environment data and applies them to the on-board computer 11, which stores these data. The data of the operating state and/or operation environment are applied to
10 the interface 10, which further conveys these data to a function unit 12, which adapts them to the respectively detected operating state or detected operation environment, to adapt the acoustic references 8. Basically, the interface 10 can also be coupled without an intermediate circuit of an on-board computer to the detector 13 (connection 12b). A detected operating state would be, for example, the operating state of a blower or also the respective speed of the
15 motorcar. The operation environment data could indicate, for example, rainy weather or also the actual condition of the road on which the motorcar is driving.

Preferably, the described system can generate speech pause models with suitable acoustic references 8. For speech pauses an acoustic signal received from the microphone 2 only contains noise signal portions, but no speech signal portions by which the
20 devices/function units 1a or 1b are to be controlled.

An embodiment of the invention comprises, in dependence on a detected operating state or a detected operation environment, confining the vocabulary of the speech recognition system 3, which is combined by the lexicon 9 to a sub-set of words which are rendered available as effective speech control signals (function block 13). With this
25 arrangement the calculation operations necessary for the comparison procedures of the function block 6 are reduced.

Fig. 2 shows an example of generating an acoustic reference 8a from the set of acoustic references 8. By the function unit 12 the basic reference, which had already been assigned to the operating state or operation environment before the speech recognition system
30 3 was taken into operation, and which best corresponds to the actually detected operating state or operation environment, respectively, is selected from a set of a priori given and predefined basic references 20-1, 20-2 to 20-n for speech pause sections. The selection of a basic reference is symbolically shown by a switch 21. Function block 22 combines an optional adaptation of the selected basic reference to reach the more accurate modeling of the

detected operating state or operation environment, respectively, to thus form the acoustic reference 8a to be used for the respective speech pause section. If, for example, an acoustic basic reference corresponds to a noise signal portion that is derived from rain noise, during the adaptation in block 22 an adaptation will take place to the detected strength of the rain, where the strength of the rain corresponds to a respective interference/noise signal level in the motorcar.

Fig. 3 shows a further variant for the generation of the acoustic reference 8a for a speech pause section. Just like fig. 2, already a priori predefined basic references are provided for speech pause sections (blocks 30-1, 30-2 to 30-n) by means of which the acoustic reference 8a is formed. In the embodiment shown in Fig. 3, however, not a single basic reference is selected. On the contrary, all the basic references are applied to a function unit 31 in which first a weighting and, as the case may be, also an adaptation of the basic references is performed (blocks 32-1, 32-2 to 32-n) in dependence on the respectively detected operating state or operation environment. The basic references weighted/adapted in this manner are finally combined to a single acoustic reference in a unit 33, which is the acoustic reference 8a to be used for the speech pause section under consideration.

The invention is not restricted to speech pause modeling. Basically, also the word sub-units can be adapted to respective acoustic references 8 in like manner to a detected operating state or operation environment of the motorcar. The acoustic reference 8a would then form the base for the adaptation of acoustic references 8 that represent word sub-sections, to model noise signal portions of an acoustic signal captured by the microphone 2.

In addition, the invention described is not restricted to the use in motorcars. The invention is basically applicable to all speech-controlled devices in which noise signals are superimposed on speech control signals, which noise signals can be indirectly determined by the detection of the operating state or operation environment of such a device.